Available online at www.sciencedirect.com
SCIENCE DIRECT®

Journal of Human Evolution 44 (2003) 665–683

Bias and error in estimates of mean shape in geometric morphometrics

F. James Rohlf

Department of Ecology and Evolution, State University of New York, Stony Brook, New York 11794-5245, USA

Received 5 November 2002; accepted 8 March 2003

Abstract

Sampling experiments were performed to investigate mean square error and bias in estimates of mean shape produced by different geometric morphometric methods. The experiments use the isotropic error model, which assumes equal and independent variation at each landmark. The case of three landmarks in the plane (i.e., triangles) was emphasized because it could be investigated systematically and the results displayed on the printed page. The amount of error in the estimates was displayed as RMSE surfaces over the space of all possible configurations of three landmarks. Patterns of bias were shown as vector fields over this same space. Experiments were also performed using particular combinations of four or more landmarks in both two and three dimensions.

It was found that the generalized Procrustes analysis method produced estimates with the least error and no pattern of bias. Averages of Bookstein shape coordinates performed well if the longest edge was used as the baseline. The method of moments (Stoyan, 1990, Model. Biomet. J. 32, 843) used in EDMA (Lele, 1993, Math. Geol. 25, 573) exhibits larger errors. When variation is not small, it also shows a pattern of bias for isosceles triangles with one side much shorter than the other two and for triangles whose vertices are approximately collinear causing them to resemble their own reflections. Similar problems were found for the log-distance method of Rao and Suryawanshi (1996, Proc. Nat. Acad. Sci. 95, 4121). These results and their implications for the application of different geometric morphometric methods are discussed.

Keywords: Shape coordinates; EDMA; Tangent space; Procrustes; Moment estimates; Sampling experiments; Reflection invariance

Introduction

In recent years there has been an increasing interest in the use of geometric morphometric methods rather than the traditional multivariate analysis of selected distance measurements, angles, and ratios to study variation in shape. Geometric morphometric methods usually begin with digitized coordinates of a number of landmark locations. The effects of variation in the location, orientation, and scale of the specimens are eliminated and the differences that remain represent shape variation and are expressed with respect to a suite of shape variables. Statistically, the

E-mail address: rohlf@life.bio.sunysb.edu (F.J. Rohlf).

0047-2484/03/$ - see front matter
doi:10.1016/S0047-2484(03)00047-2

Table 1
Method for estimating average shape

Code	Method	Shape variables	Reflections	Reference
GPA	Average shape from a generalized Procrustes analysis	Procrustes aligned coordinates	Kept separate	Goodall (1991)
BookSC	Average of shape coordinates	Bookstein shape coordinates	Kept separate	Bookstein (1991)
Rao-d	Average of shape variables	Linear combinations of log size-scaled interlandmark distances	Not distinguished	Rao and Suryawanshi (1996)
Rao-a	Average of shape variables	Subset of interior angles	Not distinguished	Rao and Suryawanshi (1998)
MOMENT	Method of moments	Size-scaled interlandmark distances	Not distinguished	Stoyan (1990) and Lele (1993)

advantage of this approach is that with sufficient sample sizes one expects to have much higher statistical power to detect shape differences because landmark coordinates capture more information about shape than can be obtained from traditional morphometric measurements. Another advantage is that these methods usually provide better visualizations of the results. See Bookstein (1991), Dryden and Mardia (1998), and Small (1996) for general treatments of the subject and Rohlf and Marcus (1993) for an overview.

A number of geometric morphometric methods have been proposed and Table 1 lists those considered here with the codes used to identify them. A user is faced with the decision of which to use and may wonder whether the use of different methods makes any difference in practice. Rohlf (2000a) contrasted different methods in terms of their shape spaces (see statistical models section below). Rohlf (2000b) compared methods with respect to statistical power in tests for differences in mean shape between populations. The surprising complexity of the power surfaces for some of the methods was in part a consequence of the non-linearity of their shape spaces (further reasons for the complexity are provided below). The present paper, which can be viewed as a sequel to Rohlf (2000b), is concerned with two simpler and more basic properties of a morphometric method—the accuracy of the estimates of mean shape and the magnitude and pattern of any bias in the estimates.

A few definitions may be helpful in order to understand the purpose of this paper. A statistical estimator, such as a mean, is said to be *unbiased* if for all sample sizes the expected value of the estimate (the average value of the estimate) is equal to the true value of the parameter. That is, $E(\hat{\theta})=\theta$, where $\hat{\theta}$ is some estimate of the parameter θ. If not, then the estimator is said to be biased. The bias is the difference $E(\hat{\theta})-\theta$. If the estimator is only unbiased as $n\rightarrow\infty$ then the estimator is said to be *asymptotically unbiased*. Despite the negative connotation of the term, biased estimators can be useful. The magnitude of error in the estimate is usually considered more important than the mere presence of a bias. The average magnitude of error is often measured as the *mean square error*, MSE, the average squared deviation of an estimate from its true value, $MSE_{\hat{\theta}}=E(\hat{\theta}-\theta)^2=\sigma^2_{\hat{\theta}}+(E(\hat{\theta})-\theta)^2$. This statistic provides a measure of accuracy rather than just precision because it is a function of closeness of estimates to the true mean rather than just the sample mean. If, in addition to being asymptotically unbiased, a measure of variability such as the MSE of an estimator goes to zero as $n\rightarrow\infty$ then the estimator is called *consistent*. These concepts are discussed in many texts, e.g., Freeman (1963), Bancroft and Han (1981), and Knight (2000).

The purpose of the present study is to investigate the accuracy of different geometric morphometric methods in estimating the mean shape of a population and the pattern of any bias. From Mardia and Dryden (1994) we know that certain methods differ in the accuracy of their estimates and that some methods yield biased estimates of shape. However, further work was needed to determine how accuracy and bias depend on the shape being estimated as well as to extend their comparisons to other morphometric methods.

Lele (1993) raised the issue of whether different morphometric methods estimate the mean shape or form (size plus shape) consistently. He showed that the method of moments, MOMENT, used in Euclidean distance matrix analysis, EDMA, provided a consistent estimate of the mean form for the model he considered but that GPA estimates of form did not. Kent and Mardia (1997) pointed out that whether or not a morphometric method is consistent depends on the statistical model used. They showed that for homogeneous, independent variation across all landmarks (the same model used by Lele, 1993), Procrustes estimates of shape were consistent. The lack of consistency Lele reported in estimating form was in the estimates of size, not shape. Lele (1993) and Lele and Richtsmeier (2001) argue that MOMENT should be used in practical applications because it provides estimates of mean shape that are statistically consistent. While consistency is a desirable characteristic of a statistical estimator, it is an asymptotic property and thus often not of much interest when estimating parameters from small samples (Freeman, 1963). Consistency could be of interest if the MSE decreases quickly as sample size increases. This is one of the questions investigated in this study.

Rao and Suryawanshi (1996) state that there is no unique way of choosing among alternative shape functions and that inferences based on a particular choice of functions will be consistent with those based on other choices provided the probability distribution can be accurately specified. However, the alternative methods for estimation and significance testing have not been developed as Mardia and Dryden (1989) have done for Bookstein shape coordinates and thus the different methods are not equivalent in practice because different test criteria and probability distributions are used for the different methods. This means that the use of different methods can lead to different conclusions—especially when shape variation is not small. The results presented below are based on sampling experiments using Monte Carlo simulations as in the studies by Coward and Conathy (1996), Lele and Cole (1996), and Rohlf (2000b). However, those studies did not investigate the expected size of the error in the estimates or their bias. Mardia and Dryden (1994) investigated the bias of averages based on Bookstein shape coordinates (the BookSC method). In the present study the space of all possible configurations of $p=3$ landmarks in 2 dimensions is sampled systematically to provide the most comprehensive study of these properties to date. The case of triangular shapes is emphasized because it is possible to illustrate the results on the printed page. This is important since it makes a more intuitive understanding possible. Numerical results are given for selected examples with more than three landmarks and some generalizations are attempted. The statistical model used to model shape variation is described in the next section. The software used for these simulations (tpsBias, tpsTri, and tpsSurfPlots) are available from the Stony Brook Morphometrics website at http://life.bio.sunysb.edu/morph.

Statistical model

Goodall's (1991) perturbation model is a simple model for variation in the positions of the landmarks around their mean locations that is convenient for simulation studies such as the present one. In this model the $p \times k$ matrix of coordinates for the p landmarks (each of k dimensions) for the ith specimen is given by

$$\mathbf{X}_i = \alpha_i \boldsymbol{\mu} + \mathbf{E}_i \boldsymbol{\Omega}_i + \mathbf{1}\boldsymbol{\omega}_i^t \quad (1)$$

where α_i is a scale factor (size of the ith specimen relative to that of the average), $\boldsymbol{\mu}$ is the mean shape, $\boldsymbol{E}_i$ is a matrix of random errors (normally distributed with mean of zero), $\boldsymbol{\Omega}_i$ is a $k \times k$ matrix describing the orientation of the ith specimen (reflections excluded), $\mathbf{1}$ is a k-dimensional vector of all ones, and $\boldsymbol{\omega}_i$ is a k-dimensional vector specifying the location of the specimen in the digitizing plane (or volume). Parameters α_i, $\boldsymbol{\Omega}_i$, and $\boldsymbol{\omega}_i$ encode information unrelated to shape variation. The estimates of shape variation must be independent of effects of variation in these parameters even though a mean shape may be expressed relative to some particular coordinate system. Matrix $\mathbf{E}_i$ (when strung out as a single column vector with kp elements) has a covariance matrix, Σ. This $kp \times kp$ covariance matrix allows for covariances both within and among landmarks. In the present study only the simplest case was investigated—identical independent variation around each mean landmark position. This is the pattern of variation such as one might expect from idealized digitizing error.

The square root of the average of the squared elements of $\mathbf{E}_i$ (root mean square error, RMSE) measures the difference between an observed shape and the true mean shape after adjusting for apparent differences in coordinates due to location, orientation, and scale. In practical applications the differences in location, orientation, and scale are not known and must be estimated and only a variance can be computed rather than a mean square error. The square root of the sum of squared differences between two aligned and scaled shapes is also called a Procrustes distance (Rohlf, 1999a, calls it a Procrustes chord distance) and it can be used to measure the amount of difference between any two shapes.

Kendall (1981; 1984) showed that the space where distances between shapes correspond to Procrustes distances is a $kp-k-k(k-1)/2-1$-dimensional manifold ($2p-4$ dimensions for 2D and $3p-7$ dimensions for 3D). Intuitively, a manifold can be thought of as an extension into general dimensions of the concept of a curve or surface in two or three dimensions (Small, 1996). While a more technical definition of a manifold would be out of place here, one property needs to be mentioned. The neighborhood of each point on a manifold permits a coordinate system to describe the relative positions of points in that neighborhood (see, for example, Bishop and Goldberg, 1968). In the context of shape analysis, the set of shapes is now called Kendall's shape space. For $p=3$ landmarks in $k=2$ dimensions (triangles in the plane) the manifold corresponds to the surface of an ordinary sphere with radius 1/2 (a 2-dimensional curved space). With the exception of the degenerate shape consisting of p coincident landmarks, every possible shape maps to a unique position in Kendall shape space. In practical applications, a generalized Procrustes analysis (GPA) is performed to estimate a mean shape and to align the specimens to it (Rohlf and Slice, 1990). Rohlf (1999) and Slice (2001) showed that after a GPA the points corresponding to the aligned shapes can be represented as points on the surface of a hemisphere. Points corresponding to these aligned specimens can be orthogonally projected onto a linear space that is also tangent to Kendall's shape space. This is possible because for small variation in shape the points will be in the same neighborhood and a Cartesian coordinate system can be used to describe the relative positions of the points. This approximating space is called Kendall tangent space by Kent (1994) (however, subsequent authors often refer to it as Kent tangent space). Because the tangent space is linear, one can use standard linear multivariate statistical methods. This space is illustrated in Fig. 1 for configurations of three landmarks. Sample configurations of three landmarks are shown as triangles at positions corresponding to their positions in this tangent space. In this illustration, an equilateral triangle has been arbitrarily selected as the reference point at the origin of this space. In practical applications the mean shape would be used. It corresponds to the point of tangency between this tangent space and Kendall's shape space. This space has a radius of 1. See Rohlf (1999b) and Slice (2001) for more information. The mapping of all possible triangles to unique positions in this space allows one to plot various properties of triangles as surfaces lofted above this space. This was done for statistical power in Rohlf (2000b). In the present study, this will be done for average error in estimates of the mean shape. In addition, the pattern and magnitude of bias for a particular method will be shown as a vector field superimposed on this space.

Shape spaces corresponding to the other morphometric methods are very different. Rohlf (2000a) illustrates these shape spaces. Readers may find the tpsTri software (Rohlf, 2002) useful for visualizing the statistical distribution of triangles using various morphometric methods. An important way in which the distance and angle-based methods differ from Procrustes-based methods is that the former are invariant to reflections of the

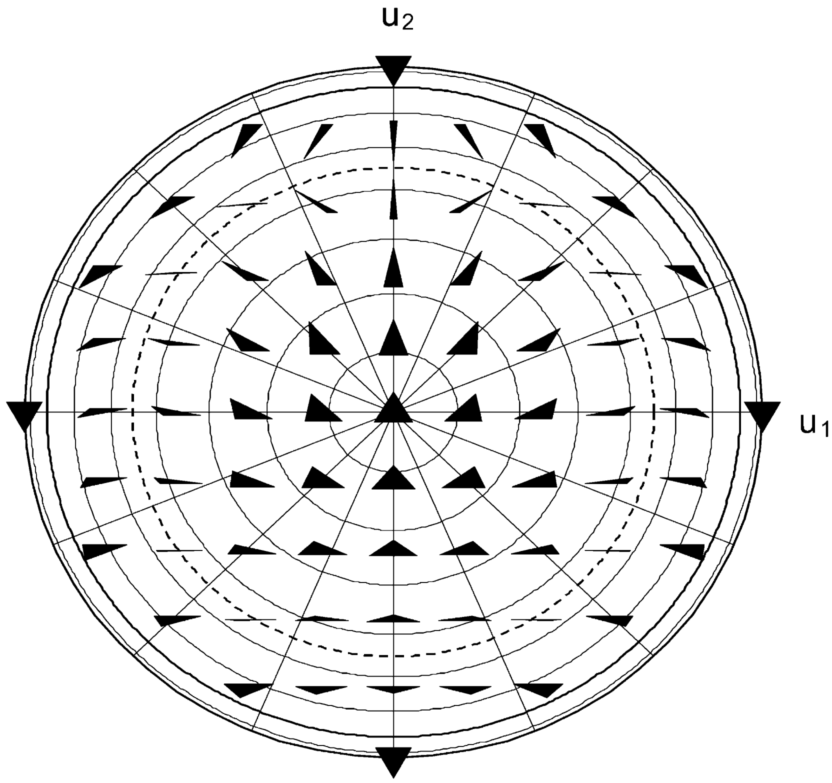

Fig. 1. Kendall tangent space for triangles using an equilateral triangle as the reference shape. The u_1,u_2 axes correspond to the estimate of the uniform component given in Rohlf (1996). Each triangle is centered on its u_1,u_2 coordinates, which range from −1 to 1. Points outside of the dashed circle correspond to triangles that are reflections of those within. All points on the outer circle correspond to the same triangle—the reflection of the equilateral triangle used as the reference. Exemplar triangles drawn using the AB edge as a baseline and landmark C as the free vertex. This figure was prepared using the tpsTri software.

configurations of landmarks. Thus, in the shape spaces for these methods, shapes shown in the region beyond the dashed circle in Fig. 1 are mapped back on top of those within that circle (see Rohlf, 2000a). The present study will show that the inability to distinguish real shape differences that appear to be reflections can be an important source of error and bias in these methods.

Estimates of average shape

Several procedures have been proposed to estimate the mean shape for a population. Table 1 lists those that will be investigated in this paper and gives the codes that will be used to identify them. A summary of the methods is given below.

Generalized procrustes analysis estimates

This approach defines the average shape as the configuration of points with minimal sum of squared Procrustes distances to every shape in the sample. It is usually computed using an iterative procedure. There are variants of this method depending on how the Procrustes distance is defined. In the type of partial Procrustes analysis used here, the aligned configurations are constrained to have unit centroid size. The method is described in Rohlf and Slice (1990) and in Slice (2001).

Average of bookstein shape coordinates

Bookstein (1991) shape coordinates for a triangle (also called the two-point or edge registration method) can be expressed in several ways. The shape coordinates for a triangle are simply the coordinates of a free vertex, say C, after the triangle has been translated, rotated, and scaled so that one vertex, A, is at the origin (0,0) and another vertex, B, is at (1,0). A convenient equation is

$$\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \frac{1}{\|B-A\|^2} \begin{pmatrix} x_b - x_a & y_b - y_a \\ -(y_b - y_a) & x_b - x_a \end{pmatrix} \begin{pmatrix} x_c - x_a \\ y_c - y_a \end{pmatrix}, \quad (2)$$

where $\|B-A\|^2 = (x_b - x_a)^2 + (y_b - y_a)^2$, is the squared length of the baseline. Some authors (e.g., Dryden and Mardia, 1998) subtract 1/2 from v_1 which corresponds to placing vertices A and B at $(-1/2,0)$ and $(1/2,0)$, respectively. Any additional landmarks can be substituted for vertex C in the above equation to yield additional shape coordinates. Problems are to be expected using this method if for some specimens the chosen baseline is not the longest edge (Rohlf, 2000a).

The average of the shape coordinates in a sample can be used as an estimate of the mean shape. Bookstein (1986), Kent and Mardia (1997), and Lele (1993) point out that estimates using this method are biased. Mardia and Dryden (1994; p. 366) prove that the average of Bookstein shape coordinates is biased towards zero and suggest dividing the estimated shape coordinates by the expression

$$1 - e^{-1/4\tau^2} \quad (3)$$

to remove bias, where $\tau = \frac{\sigma}{\delta_{12}}$, σ is the standard deviation of the error at each landmark, and δ_{12} is the true length of the baseline. The use of this correction requires knowledge of the true length of the baseline. Mardia and Dryden (1994) refer to τ as a coefficient of shape variation.

Average of Rao and Suryawanshi (1996) shape variables

Rao and Suryawanshi (1996) proposed comparing samples of shapes using a shape distance based on the logs of the interlandmark distances. Specifically, they used

$$\mathbf{d}^{(s)} = \mathbf{Hd} \quad (4)$$

as the set of $m-1$ shape variables, where $\mathbf{d}$ is the vector of logs of the $m = p-1p/2$ distances between all pairs of landmarks, $\mathbf{H}$ is an $m-1 \times m$ matrix of rank $m-1$ such that $\mathbf{H1} = \mathbf{0}$ (a Helmert matrix with the first row deleted), and $\mathbf{1}$ is a vector with all entries equal to 1. This equation projects a vector of log-distances onto a space orthogonal to their mean (thus size-scaling it by removing the log of the geometric mean interlandmark distances). The matrix $\mathbf{H}$ is not unique but its rows are orthogonal, so different choices simply correspond to rotations of the space (which have no effect on distances between shapes or on the proposed statistical tests). The average of the $\mathbf{d}^{(s)}$ are used as an estimate of the mean shape.

Average of Rao and Suryawanshi (1998) shape variables

Rao and Suryawanshi (1998) suggested that a natural method of comparing shapes is to compare angles from a triangulation of landmarks. Since the three angles in a triangle sum to a constant (π radians), only two arbitrarily selected angles from each triangle are used as shape variables. The mean shape of a triangle can be estimated using averages of these angles. In this study I used the angles at either end of the baseline used for shape coordinates. Mardia and Dryden (1994) refer to this method as the “naïve angle estimator”. Dryden and Mardia (1998, pp. 26-27) discuss the use of angles and point out that there is a loss of information and that there are pathological cases.

While several transformations of the shape variables were suggested by Rao and Suryawanshi (1998) to reduce departures from normality, they were not used in this study since their examples were based on the direct use of angles.

Method of moments

Stoyan (1990) and Lele (1993) suggested the method of moments to estimate an average form (size + shape). First, the average squared distances between each pair of landmarks are corrected for bias. For 2D shapes the distances between pairs of landmarks in the average form are estimated as

$$\sqrt[4]{\bar{e}_{ij}^2 - s_{e_{ij}}^2}, \qquad (5)$$

where $\bar{e}_{ij}$ is the average squared distance between landmarks i and j, and $s_{e_{ij}}^2$ is the variance of the squared interlandmark distances between landmarks i and j. For 3D forms the average is estimated as

$$\sqrt[4]{\bar{e}_{ij}^2 - \frac{3}{2} s_{e_{ij}}^2}. \qquad (6)$$

Second, a principal coordinates analysis (Gower, 1966) is used to compute a set of Euclidean coordinates corresponding to this matrix of corrected distances between all pairs of landmarks. This matrix may *not* correspond to a valid distance matrix (Lele and Richtsmeier, 2001) for p points in a k-dimensional Euclidean space ($p>k$). In that case only the first k eigenvectors (corresponding to positive eigenvalues) are used.

Sampling experiments

All previous studies (except Rohlf, 2000b) performed sampling experiments (Monte Carlo simulations) using just a few selected shapes. In this study a more systematic approach is used. In the first set of experiments, 481 different configurations of three landmarks were selected systematically from the space of all possible configurations and each was used as the mean of a population of shapes. Fig. 1 shows the pattern of sampling (only a subset of the triangles is shown). However, only 217 shapes were used for methods that ignore reflections. Only shapes within the dashed circle in Fig. 1 were used in this case. The triangle that would lie exactly on the outer boundary of Fig. 1 (the maximally different shape, which in this case is a reflection of the shape at the center) was not used because its location is not uniquely defined in this projection. In the second set of experiments, shapes were selected systematically along transects through shape space. Experiments were also performed using particular configurations of landmarks to either confirm results of previous authors or else to explore shape space for more than just three landmarks. In these experiments 5,000 random samples of size $n=20$ were used for each population except for the experiments on the effect of varying sample sizes.

For each mean shape investigated, random samples were constructed by adding random normal deviates with a mean of zero and a specified standard deviation σ to each of the coordinates of each mean shape. In most cases the mean shapes were adjusted to have a centroid size of 1 before adding the random deviates. The particular combinations of n and σ used and whether or not scaling was performed before the addition of random deviates are given in the figure legends and described below along with the results of each experiment.

The morphometric methods described in the previous section were used to estimate the mean shape in each of the samples. There are a number of ways in which the sample means could be compared to the true means. In order to compare the different methods the same technique must be used across all methods. Based on the results of Mardia and Dryden (1994) and Rohlf (2000a), Procrustes methods were used (see the discussion for the effect of using an alternative method). The sample means were Procrustes superimposed on the true population means and the square root of the sum of squared differences in coordinates (a root mean square error, RMSE) in the landmark coordinates was recorded as a measure of how close the estimates were to their true values. The sample means were also averaged to detect whether there was a systematic error (bias) in how they differed from the true mean. The method of averaging the sample means differed somewhat for the different morphometric methods being tested. The shape variables were averaged for the BookSC, Rao-d, and Rao-a methods. The averages of the Procrustes aligned coordinates were used to average the sample means from the GPA and EDMA methods. Because so many (5,000) means were being averaged for each case being considered, different methods of averaging would be expected to give very similar results. These overall mean triangles were then compared to the known population means.

Results

Fig. 2 shows RMSE, our measure of the magnitude of the differences between estimates and the true mean shapes, as a surface lofted above Kendall's tangent space for various estimates of the mean shape. Besides being as low as possible, the surface should also be flat, indicating that the error in estimation of the mean shape does not depend on the shape being estimated. Only the surface for the GPA method (Fig. 2A) was flat. The surface corresponding to the BookSC method (Fig. 2B) shows, as expected, much larger error when the baseline is small (u_1 near 0 and u_2 near $\sqrt{0.5}$ in Fig. 1). Because the other methods ignore reflections, their surfaces only correspond to the region within the dashed circle in Fig. 1. These other methods have larger errors for shapes that are near the outer boundary of their space. The outer boundary corresponds to flat triangles in which the landmarks are collinear. The parts of the RMSE surfaces with highest error in Fig. 2C and Fig. 2E correspond to the three possible types of isosceles triangles with one side much shorter than the other two. Because it is difficult to judge the relative heights of surfaces from these illustrations, Fig. 3 and Fig. 4 give cross-sections through these surfaces. The GPA method generally has the least error. The error for the BookSC method is only slightly larger except, as expected, when the baseline is relatively short (corresponding to shapes near the top of the u_2 axis in Fig. 1). Note how the error in the BookSC method is virtually identical to that of the GPA method at the left of Fig. 4. The error in the other methods is much higher except for shapes close to the origin in Fig. 1.

The possibility of bias in the estimates of shape was investigated by comparing the average of the 5,000 shapes estimated above for each population with the true shapes from which the samples were generated. The results are illustrated in Fig. 5 as vectors going from the positions corresponding to the true shapes to those of the average of their sample estimates. The lengths of the vectors in subfigures A to E have been exaggerated by a factor of 2 in order to make them easier to see. For the GPA method (Fig. 5A) the bias vectors were very short and there was no discernable pattern. In contrast, the BookSC shows a pattern of strong bias away from the region corresponding to a very short baseline. The bias correction factor of Mardia and Dryden (1994) was found to work well when, as in a simulation, the true lengths of the baselines are known. Its use was also found to reduce the RMSE of the estimates. Unfortunately, it gives very poor results (not shown) when sample estimates of the baseline length are used. The Rao-d and MOMENT methods (Fig. 5C and E) show strong biases for shapes corresponding to triangles where the vertices are close to being collinear and one edge is much shorter than the other two (in contrast to collinear triangles in which one edge is much longer than the other two). The estimates based on the Rao-a method (Fig. 5D) show a strong bias for any shape where the vertices are close to being collinear.

Actually, the pattern of bias is more complex than described above. Fig. 5F shows the vectors for the MOMENT method drawn to the same length so that the directions of the shorter vectors can be seen. Not much confidence should be given to the directions of the vectors near the origin (it would take much larger sample sizes to reliably estimate the directions of very short vectors). The other vectors point away from the origin—except for those near the outer limit, which point inwards.

The results given above were all based on samples of $n = 20$ shapes. Fig. 6 and Fig. 7 show the effects of increasing sample size on the RMSE for an oblique and an isosceles triangle (shown on the figures). In both cases the GPA method uniformly has the smallest error. For the oblique triangle the MOMENT, Rao-d, and BookSC methods have much larger errors than GPA for the smaller sample sizes but the difference rapidly decreases as sample sizes increase. The error in the Rao-a method decreases more slowly. While the differences are less, the relative rankings of the methods are the same even for sample sizes of $n = 10{,}000$ (not shown). The results for the isosceles triangle (Fig. 7) are rather different. The GPA and BookSC methods yield very similar and substantially smaller RMSE than the other methods. While the error for the MOMENT method is much higher for small sample sizes, it

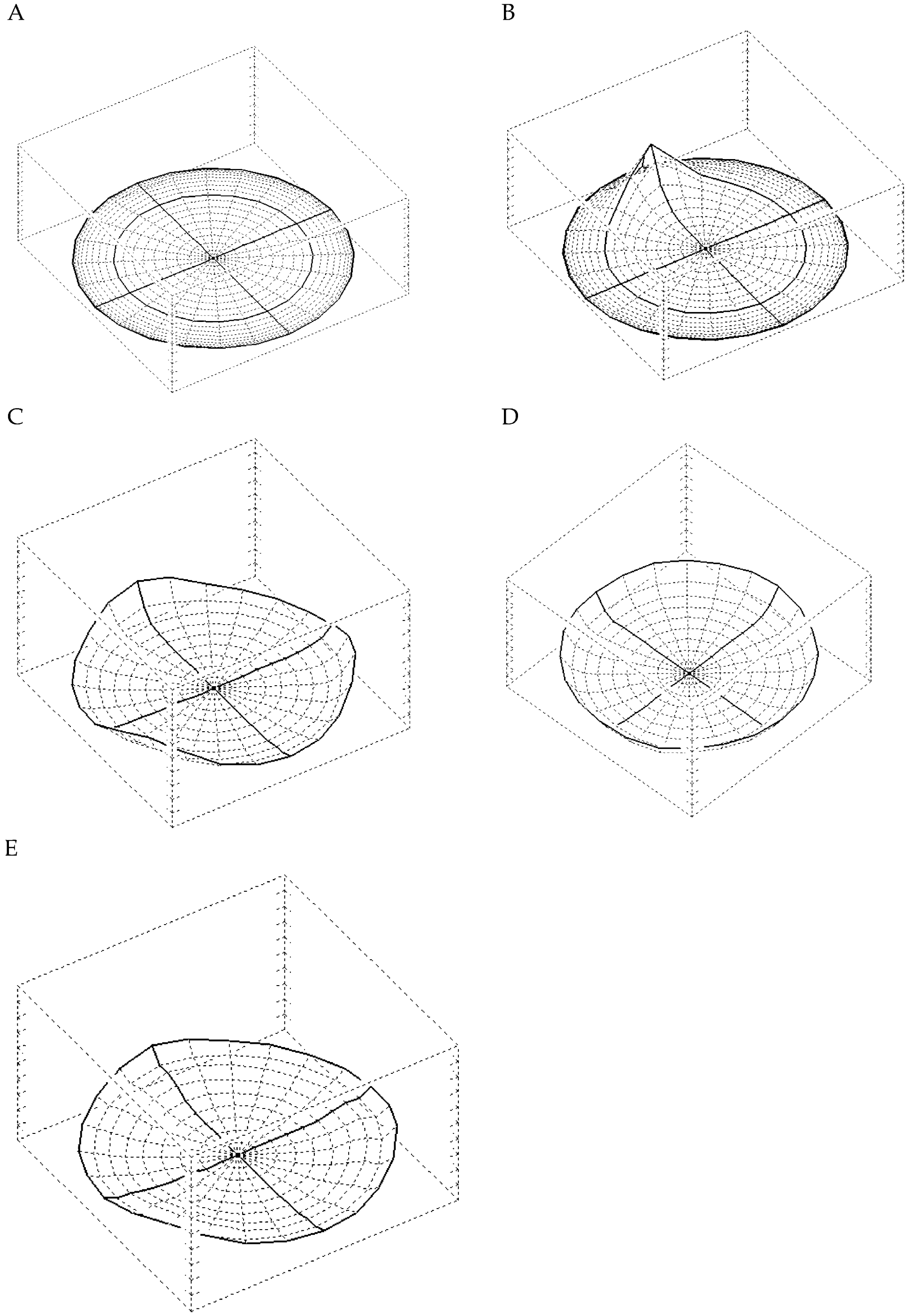

Fig. 2. Root mean square error surfaces. (A) GPA, (B) BookSC, (C) Rao-d, (D) Rao-a, (E) MOMENT. Each point is based on 5,000 samples of size $n = 20$ and normal deviates with $\sigma = 0.15$ added to coordinates of triangles scaled to unit centroid size. Note that in panels C to E (corresponding to methods that ignore reflections) the outer curve corresponds to the bold inner circle on panels A and B. These plots were prepared using the tpsSurfPlots software.

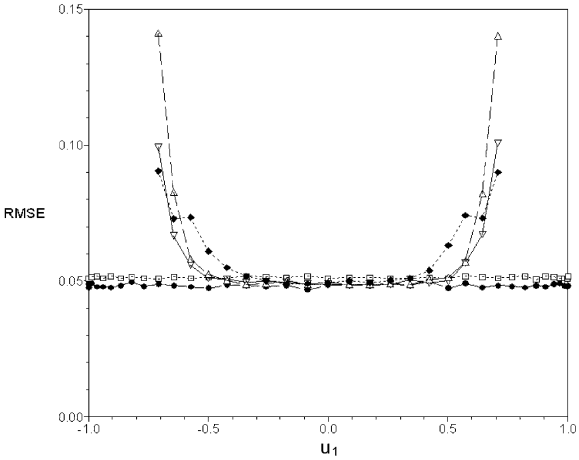

Fig. 3. RMSE curves from Fig. 2 as a function of position along the u_1 (horizontal) axis of Fig. 1. Each point is based on 5,000 samples of size $n = 20$ and normal deviates with $\sigma = 0.15$ added to coordinates of triangles scaled to unit centroid size. GPA = ●, BookSC = □, Rao-d = ▽, Rao-a = △, MOMENT = ◆.

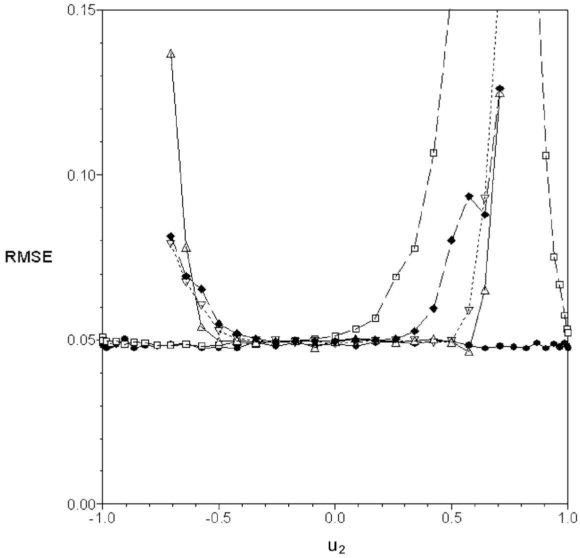

Fig. 4. RMSE curves from Fig. 2 as a function of position along the u_2 (vertical) axis of Fig. 1. Each point is based on 5,000 samples of size $n = 20$ and normal deviates with $\sigma = 0.15$ added to coordinates of triangles scaled to unit centroid size. GPA = ●, BookSC = □, Rao-d = ▽, Rao-a = △, MOMENT = ◆.

decreases rapidly. The RMSE for the Rao-d and Rao-a methods decrease very slowly.

Another sampling experiment was performed to investigate why the direction of the bias for MOMENT and Rao-d methods differed depending on the location of a point in shape space. A vertical transect across Fig. 1 seemed to capture many cases of interest. For each triangle, v_1 (the X-coordinate of Bookstein shape coordinates), was fixed at 0.5 and v_2 (the Y-coordinate) varied. In Fig. 8A the relationship is shown for the interval for v_2 going from 0 (a flat isosceles triangle) to 4 (an isosceles triangle with a baseline 1/4 of its height). Fig. 8B shows the detail for v_2 in the interval from 0 to 1. The Rao-d and MOMENT curves intersect at $v_2=\sqrt{3/4}$, which is the point corresponding to an equilateral triangle. These results show that for values of v_2 less than about 0.2 the shapes estimated by MOMENT were closer to an equilateral triangle than the true mean triangles and for larger values of v_2 the estimated triangles were flatter than the true mean triangles. The estimates for Rao-d were always closer to an equilateral triangle than the true mean triangle. The results for the GPA method are not shown because they differed so little from the true mean triangles represented by the diagonal line.

It is difficult to systematically investigate all the possibilities when there are more than three landmarks. For that reason only a few experiments and tests of special cases are reported here. The first experiment was designed to investigate the properties of the simplest case for 3-dimensional data. An experiment analogous to that described above for triangles in the plane was made to study whether there were analogous patterns of bias. In this experiment, the first three 3-dimensional landmarks had coordinates $(\frac{1}{3}\sqrt{3},0,0)$, $(-\frac{1}{6}\sqrt{3},1/2,0)$, and $(-\frac{1}{6}\sqrt{3},-1/2,0)$, corresponding to an equilateral triangle in the x, y-plane. The fourth point had coordinates $(0,0,\frac{1}{3}\sqrt{6}\,x)$, where x was varied from 1 (corresponding to a regular tetrahedron with all interlandmark distances equal to 1) down to zero (corresponding to a flat tetrahedron). Using $n = 20$ and $\sigma = 0.05$, the RMSE for GPA was about 0.025 over the entire range. The RMSE for MOMENT was similar for larger values of x but

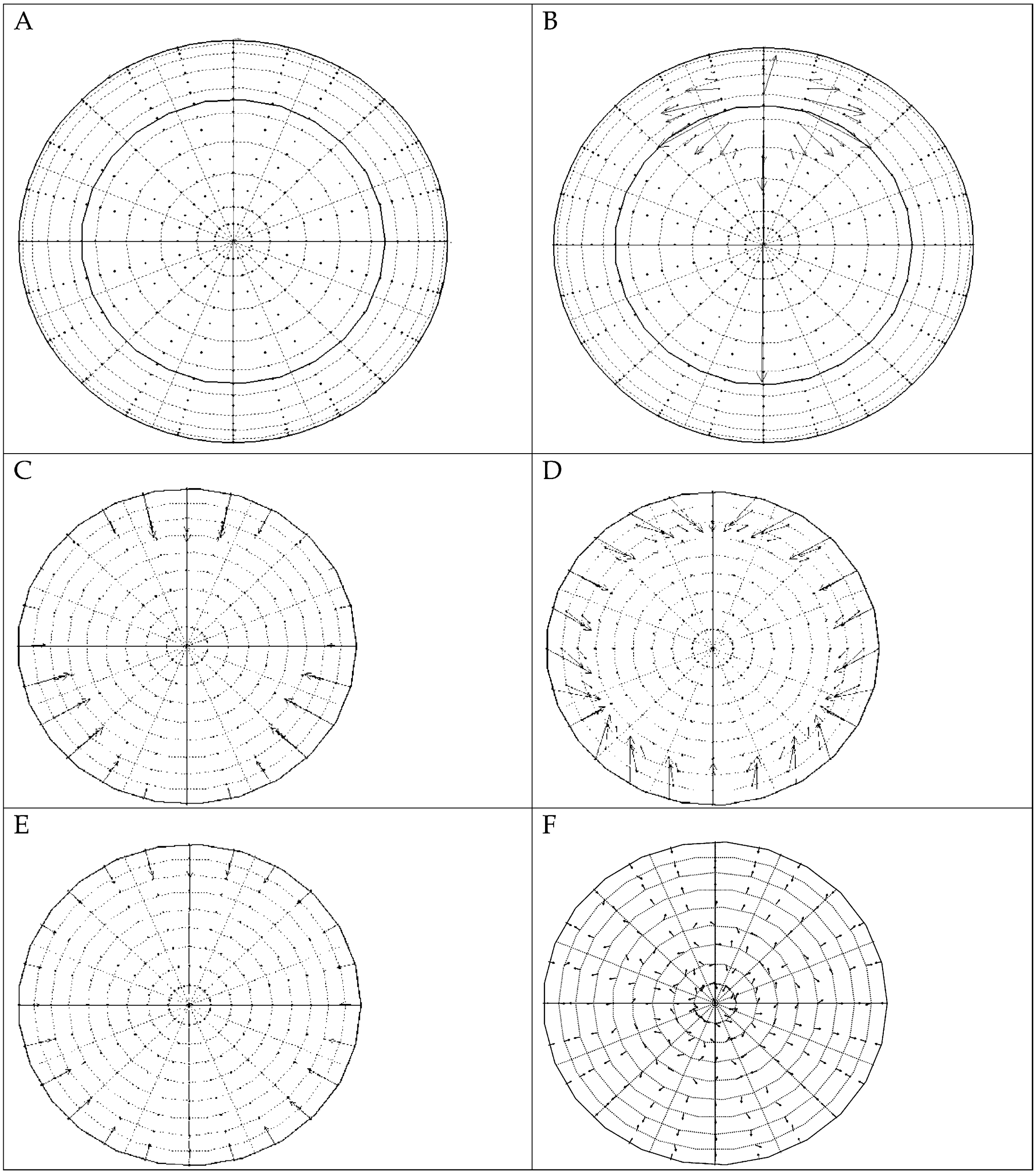

Fig. 5. Bias shown as a vector field. The vectors connect the location of a population mean and its average sample estimate. (A) GPA, (B) BookSC, (C) Rao-d, (D) Rao-a, (E) MOMENT, and (F) MOMENT showing just directions of vectors. Each point is based on 5,000 samples of size $n = 20$ and normal deviates with $\sigma = 0.15$ added to coordinates of triangles scaled to unit centroid size. Note that for panels C to F the outer circle corresponds to the bold inner circle on panels A and B. These plots were prepared using the tpsSurfPlots software.

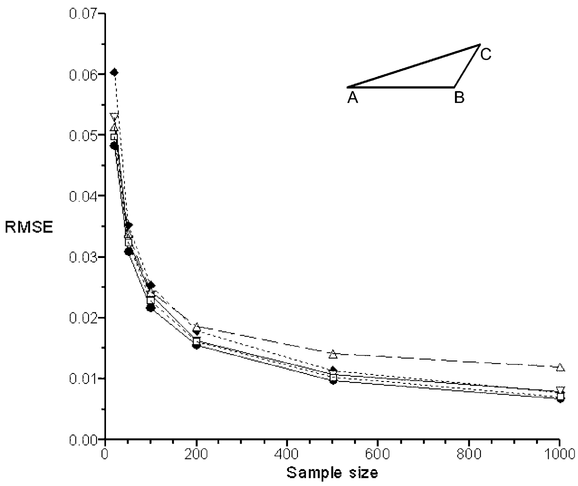

Fig. 6. RMSE as a function of sample size for an oblique triangle (see inset figure) at coordinates (0.5,0) on Fig. 1 and Bookstein shape coordinates of (1.25, 0.433). Each point is based on 5,000 samples of the specified size and normal deviates with $\sigma = 0.15$ added to coordinates of triangles scaled to unit centroid size. GPA = ●, BookSC = □, Rao-d = ▽, Rao-a = △, MOMENT = ◆.

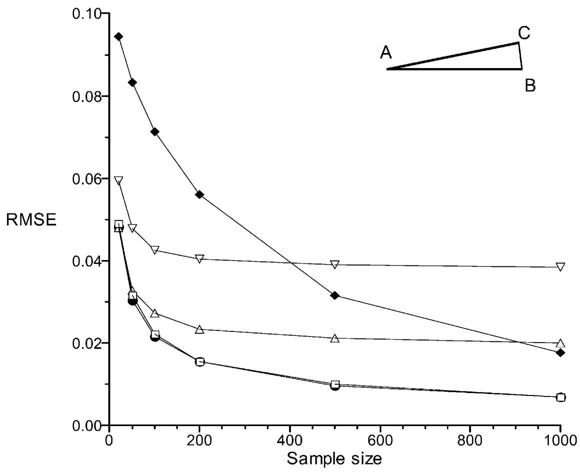

Fig. 7. RMSE as a function of sample size for an isosceles triangle (see inset figure) at coordinates (0.49709, −0.28699) on Fig. 1 and Bookstein shape coordinates of (0.9796, 0.20093). Each point is based on 5,000 samples of the specified size and normal deviates with $\sigma = 0.15$ added to coordinates of triangles scaled to unit centroid size. GPA = ●, BookSC = □, Rao-d = ▽, Rao-a = △, MOMENT = ◆.

increased for values of x less than 0.15. For example, it was 0.028 for x=0.1, 0.033 for x=0.05, and 0.048 for $x = 0.01$. This increase in RMSE was due to a bias in which the z-coordinate of the fourth landmark was estimated to be larger than it should have been. This means that for small x the estimated mean shapes were closer to a regular tetrahedron that the true shape. Unlike the case of triangles described above, there was no evidence of a bias in the opposite direction for larger values of x.

Discussion

Sampling experiments

While the σ values used in the present study were not very small, they were at the low end of the range of variation considered by Mardia and Dryden (1994). RMSE and bias are expected to be smaller in many practical applications of morphometric methods. Sampling experiments were also performed using smaller σ values. They resulted in lower RMSE values but with similar patterns of relationships among the methods as described above. When the RMSE is small, bias must also be small and one expects different methods to lead to similar estimates of the mean shape—especially when sample sizes are large (Dryden and Mardia, 1998; p. 287). However, that may not always be the case. In studies of human variation the differences that investigators try to detect are sometimes quite subtle and relatively few specimens may be available. In such studies the concerns raised here are likely to be of practical importance. When dealing with fossil material sample variability can be larger due to problems in preservation and of making measurements on such material. Kent and Mardia (1997) state that in typical applications the interest is often in testing for differences in shape rather than the isolated estimation of an average shape and that in performing such tests bias calculations are of less interest because they tend to cancel out when similar shapes are being compared. However, an examination of Figs. 5 and 8 shows that this need not be the case for all methods. Once a difference in shape has been established it is natural to want to describe what the differences are and for that description one prefers unbiased

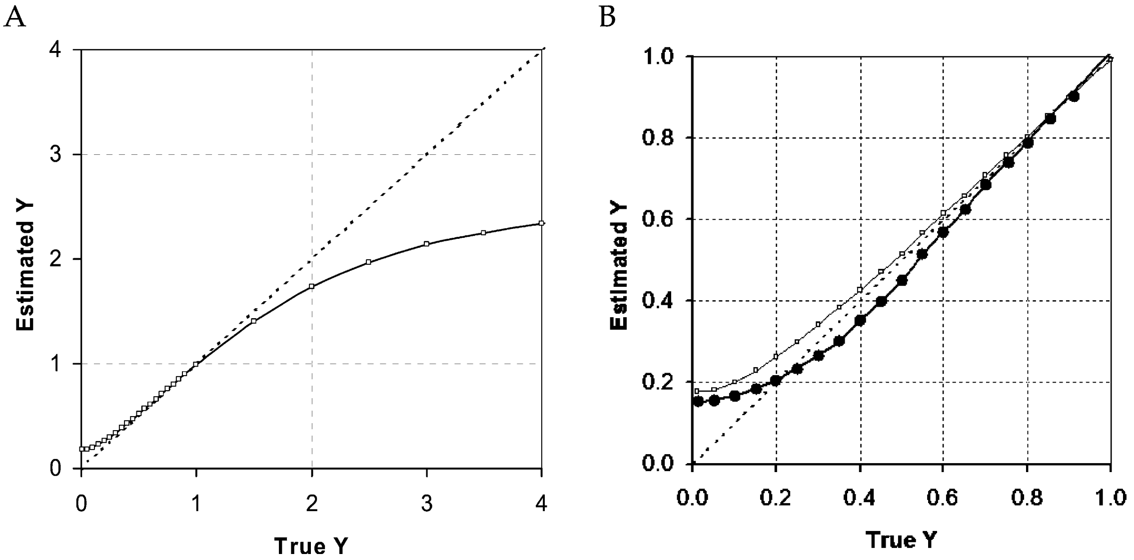

Fig. 8. Results of sampling experiments for triangles with the X-coordinate of Bookstein shape coordinates fixed at 0.5 and the Y-coordinate varying. Each point is based on 5,000 samples of size $n = 20$ and normal deviates with $\sigma = 0.3$ added to the coordinates of a triangle with specified Bookstein shape coordinates. (A) For the height of the triangle in the interval 0 to 4. (B) Showing detail within the interval 0 to 1. MOMENT = ●, Rao-d = □. Dotted line corresponds to the identity relationship. GPA (not shown) follows the identity line closely.

estimates of the mean shape with minimal RMSE.

In sampling experiments, such as used in the present study, there is always a concern that results could be spurious due to programming errors (as in Lele and Cole, 1996). To guard against that, results for selected shapes were compared against those obtained from programs developed using MATLAB (MathWorks, 2000). The MATLAB language is sufficiently different that the programs can serve as independent checks for programming errors (they cannot, of course, be used as checks of conceptual errors because they were written by the same person). The results were also compared to the values for an equilateral triangle and an isosceles triangle given in Stoyan and Frenz (1993) and for results in Mardia and Dryden (1994). For the isosceles triangle, the software used here duplicated closely the results reported by Mardia and Dryden (1994) for Bookstein shape coordinates and for GPA with scaling. The results for Rao-a (which they call the naïve angle estimate) did not match very well. For $\tau = 0.3$ and 0.6 they report a negative bias for both v_1 and v_2 whereas the bias in v_2 was found here to be positive. There were also large disagreements in the MOMENT estimates. These seem to be due to the fact that, for large values of τ, equation (5) is often undefined because the radican (the quantity under the radical sign) is negative. Their results were duplicated closely (both for the estimates and their standard deviations) by using the absolute value of the radican rather than by replacing it with a zero. A value of zero is more appropriate because a negative radican represents strong evidence that the distance is small. A value of zero was used in the present study. The results reported by Stoyan and Frenz (1993) for the MOMENT method do not match either result very closely (their bias estimates were of opposite signs and had much larger standard deviations). The results they reported for the equilateral triangle did not match very well either. The reason for these differences is unclear.

Comparison of methods

It could be argued that RMSE should not be used to evaluate the inaccuracy of MOMENTS

Table 2
Effect of triangle shape on bias using the EDMA1 T statistic to measure the closeness of the mean shapes to the true shapes. Values closer to 1 represent closer agreement with the true shape. Simulations based on 5000 samples of size 20. Left three columns show the effect of varying height as in Fig. 8 but using σ=0.15. Right three columns show the effect of varying the height of an isosceles triangle with a base equal to 1. Triangles not scaled to unit centroid size before adding random deviates with σ=0.3

Height$\leq\sqrt{3/4}$			*Height*$\geq\sqrt{3/4}$		
H	MOMENT	GPA	H	MOMENT	GPA
0.8660	1.0758	1.0746	0.8660	1.4879	1.1696
0.6	1.0941	1.0790	1	1.4149	1.1558
0.4	1.1213	1.1106	2	1.3619	1.1094
0.2	1.1607	1.1378	3	1.3314	1.0978
0.1	1.1666	1.1476	4	1.2389	1.0927
0.05	1.1668	1.1525	5	1.5081	1.0902
0.01	1.1664	1.1542	6	1.2827	1.0878
			7	1.2036	1.0879
			8	1.2466	1.0871
			9	1.2649	1.0852
			10	1.3504	1.0858

since it was not designed to minimize that criterion. This was investigated by using the EDMA1 T statistic (the ratio of max to min ratios of the ratios of interlandmark distances in two configurations being compared) to measure the agreement between an estimated mean shape and the true mean. The left half of Table 2 shows the results for mean triangles as used in Fig. 8. The T values for the two methods are similar but those for GPA are smaller indicating that the mean shapes estimated using GPA are closer to the true means. The results of simulations using isosceles triangles with base of length one and varying heights are given on the right half of Table 2. The T values for GPA are much smaller than those for MOMENTS. The effects of varying sample size on the triangle used in Fig. 7 are shown in Table 3 . As before, the means estimated by GPA were indicated to be closer to the true configuration than those estimated using MOMENTS. Thus, the use of RMSE in this paper to compare the closeness of the estimated means to their true values does not seem to be biased against the MOMENTS method. The second series of shapes used in Table 2 represents the least favorable case for the method of moments because one side of the triangles is much shorter than the other two (see Fig. 2E). This is consistent with the results of Rohlf (2000a) where the EDMA1 shape space was found to be most distorted for such triangles and the results of Rohlf (2000b) where the lowest power was obtained for tests involving such triangles.

Table 3
Effect of sample size on bias as in Fig. 7 but using the EDMA1 T statistic to measure the closeness of the mean shapes to the true shape. Values closer to 1 represent closer agreement with the true shape. Simulations based on 5000 samples of size 20 and 200. Triangle scaled to unit centroid size before adding random deviates with σ=0.15

	Method	
n	MOMENT	GPA
20	6.1737	1.2018
200	6.3676	1.0577

A potential problem in illustrating bias using a vector field superimposed on Kendall's tangent space is that distances and directions may not be accurately represented due to the inevitable distortion in projecting from a curved space. This should not be a problem here because the bias vectors were mostly oriented directly towards or away from the origin and distances in the tangent space are properly represented in those directions.

Lele and Richtsmeier (2001; pp. 86–87) discuss problems with the use of MOMENTS to estimate mean shape when sample sizes are small and suggest the elimination of landmarks that are close together in order to avoid equations (5) and (6) being undefined. This was a problem in some of the sampling experiments. Because shape could not be studied with fewer than three landmarks, no landmarks could be eliminated in the present study. A distance of zero was used whenever these equations were undefined. This tended to result in a degenerate or improper solution from the principal coordinates analysis used to recover landmark coordinates from the estimated average distance matrix. In such cases only the dimensions corresponding to the non-negative eigenvalues were used. Figs. 6 and 7 show that the interlandmark distance and angle-based methods are less accurate than GPA even when sample sizes are not small.

While the method of moments yields unbiased estimates of the squared distances between each of the pairs of landmarks individually, the entire set of distances taken as a whole is not an unbiased estimate of the interlandmark distances. This seems to be because the estimate does not take into account the fact that even under the isotropic error model the distances emanating from the same landmark are not independent. In fact, the set of estimated distances may not even correspond to a geometrically defined shape (i.e., the triangle inequality may be violated, Lele and Richtsmeier, 2001). When variability is not small the radican in equations (5) and (6) can be negative. In such cases the estimated distances are taken as zero because the observed average distance is too small in relation to the variance. When this happens there may be fewer than k non-negative eigenvalues (there may even be none) and the coordinates for one or more dimensions may collapse to zero. Stoyan and Frenz (1993) performed sampling experiments using an equilateral and an isosceles triangle and concluded that this approach gives results comparable to maximum likelihood. Mardia and Dryden (1994) observed that this estimator was biased and inefficient under large isotropic normal errors in moderately large samples.

The especially poor performance for the Rao-a method was not surprising. Mardia and Dryden (1994) demonstrated that it could have a large bias. The present study indicates the types of shapes for which the RMSE is expected to be large and the expected bias.

Reflection invariance

Lele and Richtsmeier (2001, pp. 92–93) discuss reflection invariance in terms of the advantage of being able to freely mix configurations of landmarks taken from either the left or right sides of an organism (often required with incomplete specimens). Such a convenience is not an issue: the problem is that real shape differences cannot be distinguished using such methods. Rohlf (1996) shows an example where a landmark is to the left of two other landmarks in one specimen but to the right of these two landmarks in another. Mathematically the shape differences in the triangles formed by these two configurations of landmarks will appear in part to be due to reflection (they are not because they are both on biologically right wings). Similar examples could be found for 3-dimensional data if, for example, one landmark varied above and below a plane defined by the three other landmarks.

A reason that the distance and angle-based methods have much larger RMSE for triangles corresponding to points close to the dashed circle in Fig. 1 seems clear. Consider a mean shape and sample variation such as shown in Fig. 9A. Because the true positions of the landmarks are close to being collinear, some sample configurations will appear to be reflections of the true shape. This is most obvious if we show the same scatter using Bookstein shape coordinates as in Fig. 9B (points below the baseline represent reflections of shapes above the baseline). Because distance and angle based methods are insensitive to reflections, one would obtain the same mean shape using these methods even if all of the points below the baseline were reflected to their corresponding positions above the baseline as in Fig. 9C. An estimate of the mean triangle based on the scatter in Fig. 9C would be much less flat than the true configuration—which was what was indicated in Fig. 5E. It is evident that it is not possible for an estimated mean triangle to be ‘exactly’ flat when using methods that are invariant to reflections unless there is no variation at the three landmarks. There will be a similar problem when estimating the mean shape for triangles in which a pair of landmarks is much closer together than the other landmarks. By ignoring these cases of apparent reflection, we lose information and as a result the estimated shapes will be less flat than they should be. These results make it easier to understand why the power surfaces for these methods (see Rohlf, 2000b) were more complex than one might have expected. The effects of ignoring apparent reflections are potentially an important problem with application of methods based on interlandmark distances and angles. In order to use them safely one must make sure these cases of apparent reflection do not occur in one’s data.

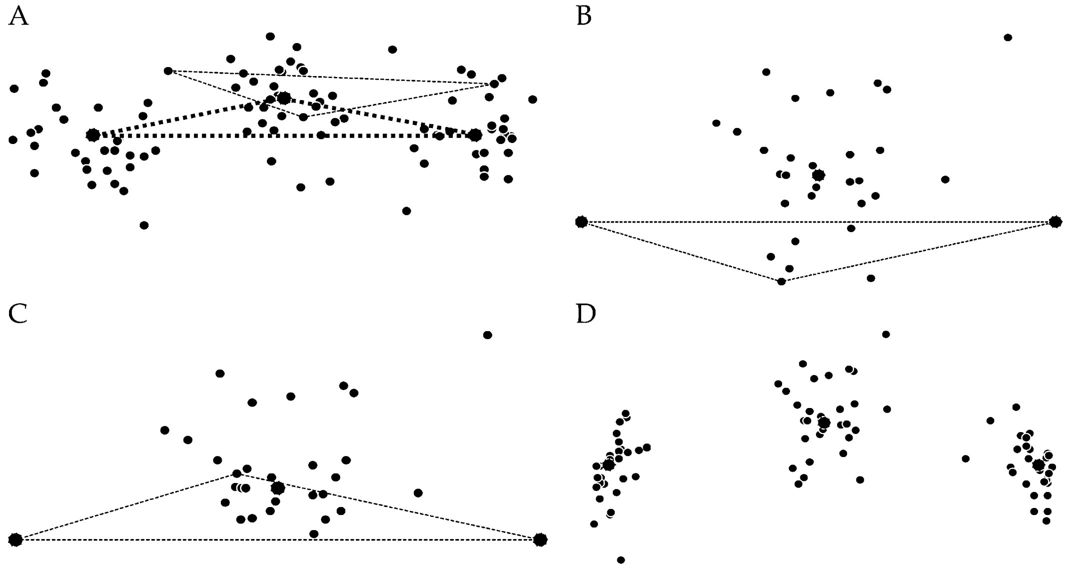

Fig. 9. An example to illustrate the problem with estimating shapes using methods that are insensitive to reflections in the data. (A) A mean shape (larger dots connected by dotted line) with sample scatter superimposed and one triangle highlighted that appears to be a reflection of the mean shape (though it is not actually a reflection because the simulated structure on which these landmarks are found was not reflected), (B) the same scatter and selected triangle shown using shape coordinates, (C) the same scatter after reflecting points below the baseline. (D) Configurations Procrustes aligned to the mean to show the distorted pattern of apparent covariance within each landmark.

The problems of insensitivity of methods based on interlandmark distances or angles to apparent reflections of configurations of landmarks can usually be avoided by adding additional landmarks but that will make it impossible to avoid having some landmarks closer together than others. That is undesirable for the distance and angle-based methods because they behave best (higher power, lower RMSE, and reduced bias) when the interlandmark distances are nearly equal. Another reflection related problem in MOMENTS is in computing the estimated landmark coordinates using a principal coordinates analysis of the estimated average distance matrix. The orientation of an eigenvector is arbitrary and thus an estimated mean configuration in one sample may be reflected relative to the configurations within the sample. When shape variation is small this situation can be recognized and corrected. This is difficult, however, when sampling variation is larger and some sample configurations appear reflected. It may not be obvious whether or not the estimated mean configuration should be reflected. This could be done in a sampling experiment by comparing the estimated configuration to the true configuration. That was not done here because it would not simulate what could be done in practice.

Limitations of the present study

The main limitation of the generality of the results from the present study is that it considers only the simplest model for shape variation—equal uncorrelated variation around each landmark. Such a model is still useful because it can serve as a standard against which to compare sample data. If a data set does not fit this model well, then that fact can be used as evidence that a more complex model should be used. It is tempting to make conclusions about the lack of fit of such a model by studying the variation around each landmark after a Procrustes superimposition of a sample onto their mean shape. Inferences about variation and covariation at each landmark is more difficult than it might seem because, as noted by Rohlf and Slice (1990: p. 57) and Slice (1993), it is possible for the superimposition itself to impose a pattern of covariation on the landmarks. Fig. 9A shows the actual simulated variation at each landmark and Fig. 9D shows the variation after a Procrustes superimposition. Simulations were performed in those studies (using the mean consensus configuration as a population mean) to make sure that the patterns being interpreted were not artifacts of the superimposition procedures. Walker (2000) documents some of the problems in estimating the pattern of variances and covariances of landmarks after a superimposition. Lele and Richtsmeier (2001; p. 94) conclude that because these estimates are poor any statistical inferences based on Procrustes estimates of variance such as confidence intervals for the mean shape or a difference between two mean shapes or principal components analysis of Procrustes residuals can be "patently misleading". These conclusions conflict with the results of Rohlf (1995), Rohlf (2000a), and Rohlf (2000b) showing that multivariate analyses based on Procrustes residuals under the isotropic error model yield correct type I error levels, have good power, and give the results one expects in a PCA. The reason for this discrepancy is that the statistical tests are not performed in the space of the organism but in shape space or its tangent approximation. Fig. 9B using Bookstein shape coordinates shows the circular scatter that one would expect for the model on which this simulation is based. Kendall tangent space coordinates would show a very similar scatter for these data since the baseline is quite long. This is an illustration of result 6.13 in Dryden and Mardia (1998) that states that if the original points have an isotropic scatter around each landmark, then the partial tangent coordinates will also be isotropic. While the proper estimation of the covariation at each landmark is an important problem that needs to be solved, it is not directly relevant for tests of mean differences in Kendall's tangent space. Lele and Richtsmeier (2001) were concerned with the wrong covariance matrix for tests of mean differences in shape. On the other hand, estimation of the relative variability at different landmarks and the patterns of covariance within and between landmarks should be of interest in many practical applications. Existing methods for estimating these variances and covariances are unsatisfactory because they can show unequal variances and strong covariances within and between landmarks when the actual variation is isotropic. The isotropic model may be more appropriate than it appears from the empirical results one obtains from the application of GPA.

Summary

Through the use of sampling experiments it was demonstrated that the generalized Procrustes estimate of the mean shape in a population was unbiased and was closer to the true shape (it had smaller mean squared error) than the other geometric morphometric methods considered here.

The results of this and related studies have important implications for selecting the morphometric methods that should be used in practical applications. If the purpose of the study is to test for differences in mean shape between samples from two or more populations then one should prefer a test that has the intended type I error rates and has the highest power (so that one has the best chance of demonstrating that there are differences between populations). Rohlf (2000b) showed that Procrustes-based tests had the highest power and that power was often much lower for methods based on interlandmark distances or angles. That study, and the one by Rohlf (1995), showed that these methods also yielded type I error rates closer to their intended levels than did the alternative methods. It is often of interest to estimate the mean shape in a sample. This may be either to summarize a sample or to be able to describe the shape differences between two or more samples. For this task it is important to use methods that yield unbiased estimates and are as close as possible to the true shape. It was shown above that Procrustes estimates show no evidence of bias and were the most accurate. If the purpose of the study is to describe the pattern of diversity of shapes in a sample then one needs to use a statistical method that does not introduce patterns of covariance into the results because of properties of its shape space. Rohlf (2000a) showed an example of how a PCA based on a Procrustes superimposition yielded the expected within-population isotropic scatter in samples taken from populations with different mean shapes. This was in contrast to the results for the Rao-d method where different within-sample covariances were obtained because of differences in the mean shapes. Similar results would be obtained using dissimilarity measures based on EDMA. Though not plotted to proper scale, Figure 6.2 of Lele and Richtsmeier (2001) shows examples of this artifact. Thus, Procrustes-based methods seem to be the most appropriate approach for all of these types of applications.

Acknowledgements

Helpful comments on an early version of the manuscript were provided by Dean C. Adams, John Kent, and Dennis E. Slice. Chris Small and Fred L. Bookstein provided help on some mathematical details. The extensive suggestions made by several anonymous reviewers and the Associate Editor were also greatly appreciated. This work was supported in part by a grant (IBN-0090445) from the Ecological and Evolutionary Physiology program of the National Science Foundation. This article is contribution no. 1110 from the Graduate Studies in Ecology and Evolution, State University of New York at Stony Brook.

References

Bancroft, T.A., Han, C.-P., 1981. Statistical theory and inference in research. Marcel Dekker, New York.

Bookstein, F.L., 1986. Size and shape spaces for landmark data in two dimensions (with discussion and rejoinder). Stat. Sci. 1, 181–242.

Bookstein, F.L., 1991. Morphometric tools for landmark data: Geometry and Biology. Cambridge University Press, New York.

Coward, W.M., Conathy, D.M., 1996. A Monte Carlo study of the inferential properties of three methods of shape comparison. Am. J. Phys. Anthropol. 99, 369–377.

Dryden, I.L., Mardia, K.V., 1998. Statistical shape analysis. John Wiley and Sons, New York.

Freeman, H., 1963. Introduction to statistical inference. Addison-Wesley, Reading.

Goodall, C.R., 1991. Procrustes methods in the statistical analysis of shape (with discussion and rejoinder). J. R. Stat. Soc. Series B 53, 285–339.

Gower, J.C., 1966. Some distance properties of latent root and vector methods used in multi-variate analysis. Biometrika 53, 325–338.

Kendall, D.G., 1981. The statistics of shape. In: Barnett, V. (Ed.), Interpreting multivariate data. Wiley, New York, pp. 75–80.

Kendall, D.G., 1984. Shape-manifolds, Procrustean metrics and complex projective spaces. Bull. London Math. Soc. 16, 81–121.

Kent, J.T., 1994. The complex Bingham distribution and shape analysis. J. Roy. Statist. Soc. B 56, 285–299.

Kent, J.T., Mardia, K.V., 1997. Consistency of Procrustes estimators. J. R. Stat. Soc. B 59, 281–290.

Knight, K., 2000. Mathematical statistics. Chapman and Hall, New York.

Lele, S., 1993. Euclidean distance matrix analysis: estimation of mean form and form difference. Math. Geol. 25, 573–602.

Lele, S., Cole, T.M. III, 1996. A new test for shape differences when variance-covariance matrices are unequal. J. Hum. Evol. 31, 193–212.

Lele, S.R., Richtsmeier, J.T., 2001. An invariant approach to statistical analysis of shapes. Chapman and Hall, New York.

Mardia, K.V., Dryden, I.L., 1989. Shape distributions for landmark data. Adv. Appl. Prob. 21, 742–755.

Mardia, K.V., Dryden, I.L., 1994. Shape averages and their bias. Adv. Applied Probability 26, 334–340.

MathWorks. 2000. MATLAB, version 6.

Rao, C.R., Suryawanshi, S., 1996. Statistical analysis of shape of objects based on landmark data. Proc. Nat. Acad. Sci. 93, 12132–12136.

Rao, C.R., Suryawanshi, S., 1998. Statistical analysis of shape through triangulation of landmarks: a study of sexual dimorphism in hominids. Proc. Nat. Acad. Sci. 95, 4121–4125.

Rohlf, F.J., 1995. Multivariate Analysis of shape using partial-warp scores. In: Mardia, K.V., Gill, C.A. (Eds.), Proceedings in Current issues in statistical shape analysis. University of Leeds, Leeds, pp. 154–158.

Rohlf, F.J., 1996. Morphometric spaces, shape components and the effects of linear transformations. In: Marcus, L.F., Corti, M., Loy, A., Naylor, G.J.P., Slice, D.E. (Eds.), Advances in morphometrics. Plenum, New York, pp. 117–129.

Rohlf, F.J., 1999. Shape statistics: Procrustes superimpositions and tangent spaces. J. Classification 16, 197–223.

Rohlf, F.J., 2000a. On the use of shape spaces to compare morphometric methods. Hystrix 11, 9–25.

Rohlf, F.J., 2000b. Statistical power comparisons among alternative morphometric methods. Am. J. Phys. Anthropol. 111, 463–478.

Rohlf, F.J., 2002. tpsTri, version 1.14. Department of Ecology and Evolution. State University of New York at Stony Brook. Available from http://life.bio.sunysb.edu/morph.

Rohlf, F.J., Marcus, L.F., 1993. A revolution in morphometrics. Trends in Ecology Evolution 8, 129–132.

Rohlf, F.J., Slice, D.E., 1990. Extensions of the Procrustes method for the optimal superimposition of landmarks. Syst. Zool. 39, 40–59.

Slice, D.E., 1993. Extensions, Comparisons, and Applications of Superimposition Methods for Morphometric Analysis. Department of Ecology and Evolution, State University of New York, Stony Brook.

Slice, D., 2001. Landmarks aligned by Procrustes analysis do not lie in Kendall's shape space. Systematic Biology 50, 141–149.

Small, C.G., 1996. The statistical theory of shape. Springer, New York.

Stoyan, D., 1990. Estimation of Distances and Variances in Booksteins Landmark Model. Biomet. J. 32, 843–849.

Stoyan, D., Frenz, M., 1993. Estimating Mean Landmark Triangles. Biomet. J. 35, 643–647.

Walker, J.A., 2000. Ability of geometric morphometric methods to estimate a known covariance matrix. Syst. Biol. 49, 686–696.